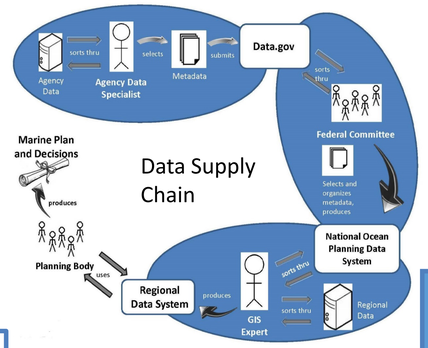
Herding cats and sorting candy: a controlled vocabulary for marine planning data

Fran Lightsom, USGS Woods Hole Coastal and Marine Science Center, flightsom@usgs.gov

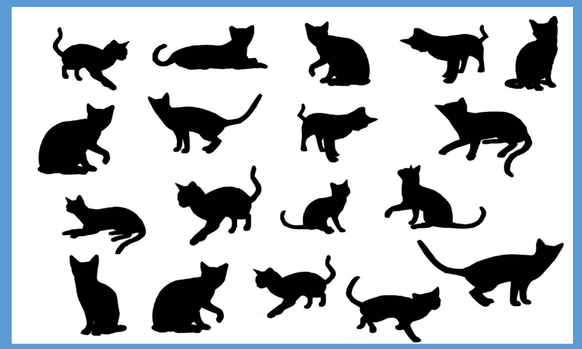
The Challenge

The **data supply chain** is a simple model of a distributed, interconnected network of independent people, organizations, and machines. The supply chain model identifies a set of data provisioning activities that are analogous to the trophic levels in the food chain. Data producers make and interpret measurements, process and integrate data, and contribute their products to data repositories, some of which are maintained by Federal Agencies. Agency Data Specialists sort through agency data holdings, select appropriate data sets, update their metadata, and submit the metadata to Data.gov. A Federal Committee sorts through Data.gov, selects appropriate metadata, and organizes it into a National Ocean Planning Data System. A Regional GIS Expert sorts through the national data system, and also through regional data holdings, to produce a Regional Data System, which is then used by the Regional Planning Body. This is simplified; multiple people, organizations, and machines do all these things.



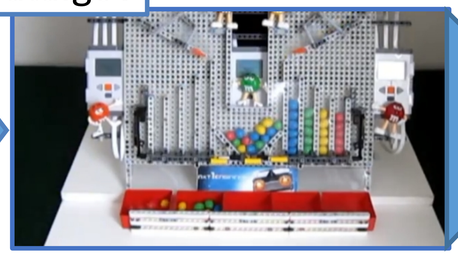
The National Ocean Policy of 2010 calls for **regional planning** bodies to plan for ecosystem-based management and coordination of a wide range of ocean and coastal activities, in order to improve the conservation of the oceans and coasts and Great Lakes. In each of 9 regions, planning bodies will represent states and territories, Federal agencies, and Indian tribes. Each of the regions is a different large marine ecosystem, with a unique combination of habitats, human uses, risks, and potential.

Meeting the Challenge

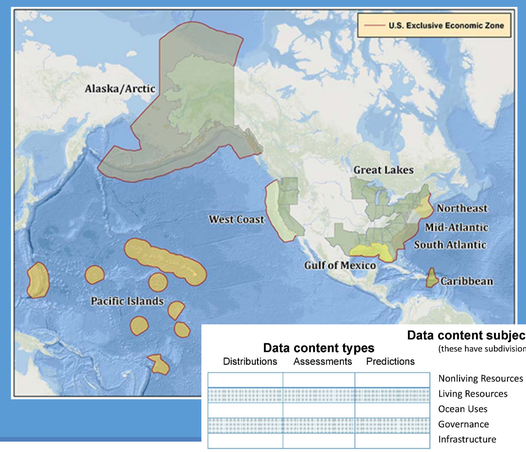


Data Supply Network

Data.gov



"M&M Dispencer" made with a Lego Mindstorms NXT kit. More information at http://www.geek.com/news/lego-mm-sorting-machine-1531882/

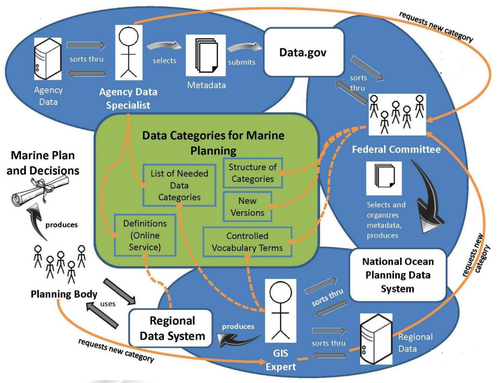


Broad, transparent foundations of high-quality data for 9 regional planning efforts.

A **controlled vocabulary** provides consistent spelling and also clear definitions to identify the broad range of data that will be needed for marine planning. The list of categories can be used as a simple checklist by people involved in sorting and selecting the data that is needed for marine planning. The controlled vocabulary adds value when the terms are also used to identify data, for example in metadata keyword fields, so that automated systems can use them to sort and search.

To be useful for **automated processing**, the categories need to be available online in a form that connects with software. Right now, the terms and definitions are only online in a PDF publication. Peter Fox and Stephan Zednik of Rensselaer Polytechnic Institute (RPI) are starting a project to convert the controlled vocabulary to linked data, put them in a vocabulary service, and use them in an application that searches the Data.gov collection to provide a customized taxonomic interface for the National Ocean Council.

The data categories for marine planning provide a **logical structure** for organizing the data. This enables users to discover what data are available, and also creates empty spaces that call attention to missing data. The categories are organized in a hierarchy of subjects, with the top level categories being nonliving resources, living resources, ocean uses, governance, and infrastructure. For each subject, we need data of three different types: distributions, assessments, and predictions. Distributions are spatial data that tell us what is where, and sometimes when, in a seasonal or cyclical way, but always talking about the present condition. Assessments are evaluations of one sort or another. The third data type is predictions, which might be observations or evaluations but are always estimates of future conditions.



Work done in partnership with Giancarlo Cicchetti (EPA Atlantic Ecology Division) and Charles Wahle (NOAA Marine Protected Areas Center).
Data categories based on a list produced by the National Ocean Council Interagency Information Management System and CMSP Data Portal Working Group .